

## 基于 MOOC 平台的国内外数据科学课程建设对比分析及经验启示

Lessons from the Comparison of Data Science in MOOC at Home and Abroad

桂思思<sup>1,2</sup> 任珂<sup>1,2</sup> 王玉琦<sup>1</sup> 程齐凯<sup>1,2</sup>

(1. 武汉大学信息管理学院, 武汉, 430072; 2. 武汉大学信息检索与知识挖掘研究所, 武汉, 430072)

**[摘要]** 基于强烈的市场需求,数据科学教育的影响正迅速扩大,与此同时,大型开放式网络课程(MOOC)为数据科学教育提供了新机遇。因此,本文在文献调研的基础上,构建了平台—组织机制—课程结构三维分析框架,基于此框架,通过网络调研法,从平台、组织机制和课程结构三个维度对比分析国内外 17 个 MOOC 平台上共计 81 门数据科学课程的开展现状。结果表明,虽然我国数据科学教育处于起步阶段,但是发展趋势较好,在平台、组织机构以及课程结构等方面仍存在改善空间。研究结果有利于丰富数据科学课程来源并建立完善的数据科学 MOOC 课程体系,从而培养社会需要的数据科学家,辅助企业和政府进行决策支持。

**[关键词]** 数据科学 MOOC 课程建设 对比分析 经验启示

**[中图分类号]** G250 **[文献标识码]** A **[文章编号]** 1003-2797(2018)03-0119-10 **DOI:** 10.13366/j.dik.2018.03.119

**[Abstract]** Due to the high demand of data scientists and the dramatic technological changes in online learning, this paper aims to improve data science education in China. Thus it presents an overview of 81 data science courses related Massive Open Online Courses (MOOC) from 17 Web sites at home and abroad by the use of a three dimensional analytical framework. Results point out that data science MOOC education in China is at the starting edge with promising development trends, which can be improved in platform, organization and courses. The results could be beneficial for the enrichment the sources of data science course and the improvement of data science course system so as to train qualified data scientists and support decision-making for government and enterprise.

**[Keywords]** Data science; Massive Open Online Course(MOOC); Course construction; Comparative analysis; Experience enlightenment

### 1 引言

自 2011 年起,大型开放式网络课程(Massive Online Open Courses, MOOC)得到了迅速发展,其具有规模化、开放性、网络化和创新性四大特征<sup>[1]</sup>,实现了优质教育资源的全球共享,学习方式、方法的创新<sup>[2]</sup>。美国教育部也认为 MOOC 平台能提供更多的学习机会、更好的授课方式、更优的学习评价体系,从而使学习者受益<sup>[3]</sup>。

数据科学(Data Science)是继云计算、大数据等概念后推出的“学科”,正成为继 OA(Open Access)、iSchool 运动之后又一影响信息科学的热流<sup>[4]</sup>。数据科学家更被评为“21 世纪最性感的职业”<sup>[5]</sup>,截至 2018 年底数据科学领域将有 14 万至 19 万个空缺岗位,且其他领域的岗位也要求掌握一定数据科学知识<sup>[6]</sup>。基于强烈的市场需求,数据科学教育的影响正迅速扩大。目前多所国、内外知名大学开设了数据科

**[作者简介]** 桂思思,博士研究生,研究方向:信息检索、查询意图、用户兴趣,Email:sgui0229@whu.edu.cn;任珂,硕士研究生,研究方向:信息检索、数据挖掘,Email:2015201040056@whu.edu.cn;王玉琦,本科生,研究方向:信息检索,Email:haitang59@foxmail.com;程齐凯,博士,讲师,研究方向:信息检索、文本挖掘,Email:chengqikai0806@163.com。

## 基于 MOOC 平台的国内外数据科学课程建设对比分析及经验启示

Lessons from the Comparison of Data Science in MOOC at Home and Abroad

桂思思 任珂 王玉琦 程齐凯

学相关课程,例如,2014年哥伦比亚大学数据科学研究所设立了数据科学硕士项目,清华大学成立了数据科学研究院并于2014年秋季开始跨学科培养硕士。然而,数据科学教育具有需求广泛、聚焦前沿热点且重视实践技能获取等特征,偏向于理论知识传授的传统教育便无法最大化地满足学习者的学习需求。在MOOC平台上开展数据科学教育可以充分利用其平台优势,增加课程的易获得性,优化教育资源分配,使教师走出传统授课模式,更加贴近前沿科技,共同为来自世界各地的学生提供学习服务和支持。

本文以国内外17个MOOC平台上81门数据科学课程为研究对象,根据MOOC系统体系结构,构建平台—组织机制—课程结构三维MOOC课程分析框架,从平台、组织机制和课程结构三个维度的10个要素,调研数据科学课程开设情况。本文的调研结果能为进一步完善数据科学MOOC教育提供一定的借鉴和参考,从而提升我国数据科学教育质量。

## 2 相关研究

目前,有关数据科学MOOC的研究主要可以分为以下五个方面:

(1)数据科学的内涵及应用研究:阐述数据科学在学术研究和实际应用中的重要作用。在学术研究中,Dhar<sup>[7]</sup>介绍了数据科学及数据科学家的实质和前景,叶鹰与马费成<sup>[8]</sup>论述了数据科学的内涵以及与信息科学的相互促进作用;在实际应用中,Hazen等<sup>[9]</sup>认为数据科学可优化数据质量,可用于供应链管理流程中的数据分析,许荻迪与江小蕾<sup>[10]</sup>阐述了数据科学对于客户分类的技术支撑和决策支持作用。

(2)数据科学线下课程体系研究:各大学开设数据科学的必要性和具体情况。王迪等<sup>[11]</sup>调研了美国22所开设数据科学的高校,认为美国数据科学学士学位教育处于初级发展阶段,并将数据科学列入硕士教学计划;苏日娜等<sup>[12]</sup>以6所iSchools高校作为调研对象,从专业学科优势、学科体系划分、课程目标、核心课程设置、课程制度等方面研究数据科学课程体系;徐昊等<sup>[13]</sup>以吉林大学为例,从课程目标和课程模块设置方面具体介绍面向通识教育的数据科学课程教学实践。

(3)利用MOOC平台开展学科教育的可行性研究。Burd等<sup>[14]</sup>认为MOOC在最大化发挥自身优势的

同时,可促进高等教育资源的合理化分配;Li与Powell<sup>[15]</sup>提出在全球教育发展大环境下,MOOC在推进高等教育开放式发展中有着极其重要的作用。除此之外,Wolff等<sup>[16]</sup>以FutureLearn平台开设的《智慧城市》课程为例,探讨信息素养在数据分析中所起的作用,Bond<sup>[17]</sup>通过文献调研法探讨MOOC开设信息素养课程的必要性及课程要求。Hardin等<sup>[18]</sup>和Liy-anaganawardena<sup>[19]</sup>均提及了MOOC平台上约翰霍普金斯大学Roger Peng教授开设的“Data Science”专项课程,但未做详细阐述。

(4)MOOC课程要素研究。分析MOOC中学习者的基本特征和课程资源特征。例如,Christensen等<sup>[20]</sup>分析了MOOC平台学习者性别、国别、年龄、受教育程度和职业等数据;Andersen等<sup>[21]</sup>分析论坛等互动系统;Chauhan与Goel<sup>[22]</sup>则从学习者角度分析了MOOC中视频,以方便不同学习者快速找到适合自己的学习视频。

(5)MOOC使用者行为研究:Mukala等<sup>[23]</sup>以课程“Process Mining: Data Science in Action”为例,分析用户的学习行为和最终成绩的相关关系;Veeramachaneni等<sup>[24]</sup>对MOOC数据科学课程的用户线上学习行为进行建模,并预测后续学习行为。

以上研究表明,学界对数据科学MOOC教育研究十分重视,但是还存在以下不足:①针对MOOC平台课程研究缺乏系统的分析框架;②对MOOC平台上开设的数据科学课程只关注用户行为分析或只进行简要介绍,没有对课程结构等核心内容进行全面系统地分析。针对这些不足,本文通过文献调研和网络调研,在前人研究成果的基础上完善MOOC平台数据科学课程研究框架,以平台、组织机构和课程结构三个维度为切入点,探讨目前MOOC数据科学课程的开设情况,以完善MOOC平台数据科学教育的方法和途径,推进数据科学教育的发展。

## 3 研究框架与研究对象

### 3.1 研究框架

基于上述文献调研可知,MOOC是以先进的平台技术为基础,以完善的组织机制为保障,以合理的课程结构为核心的开放教育系统。因此本文构建了平台—组织机制—课程结构三维MOOC课程分析框架(见图1)。

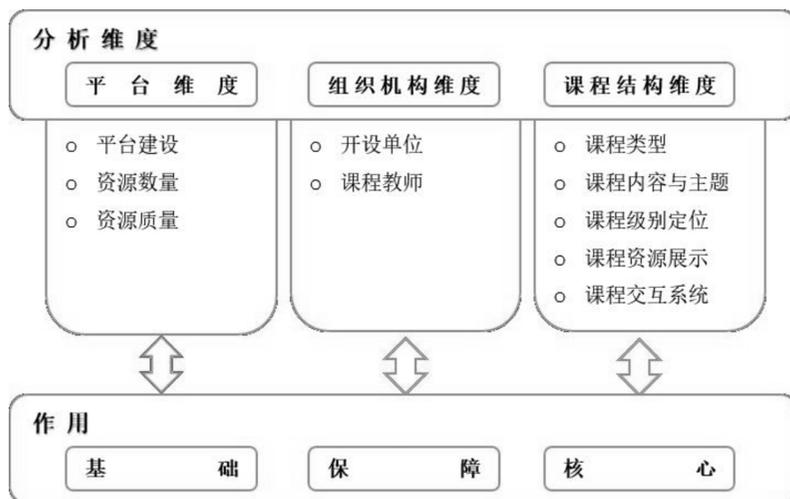


图1 平台—组织机构—课程结构三维 MOOC 课程分析框架

平台技术是 MOOC 的基础,为其所支撑的组织机构和课程结构提供运行所依赖的环境<sup>[25]</sup>;组织机构包括组织机构和合作机构,即课程所依附的教育机构以及课程提供方<sup>[26]</sup>;课程设置中除了提供视频资源、文本材料和在线答疑外,还为学习者提供各种用户交互性社区,注重对学生学习支持的服务,以及学生的学习体验<sup>[27]</sup>。

### 3.2 研究对象

MOOC 学院(或称果壳慕课网)<sup>①</sup>旨在基于真实学习者的评价,筛选和推荐最值得上的 MOOC 课程<sup>②</sup>。该网站本身不具备对 MOOC 课程资源进行运行、更新、维护、管理和推广的功能,而是作为一个搜索引擎,致力于让网民更便捷地获取课程信息,找到所求。本文不将其作为研究对象,只是根据该网站中公开课检索导航栏中机构列表,以用户能在国内顺利访问为原则<sup>③</sup>,最终确定了国内外的 17 个 MOOC 平台作为研究对象,具体如下(按拼音排序):

国外 MOOC 平台(4 个): Coursera<sup>④</sup>、edX<sup>⑤</sup>、MIT-OpenCourseWare<sup>⑥</sup>、Udacity<sup>⑦</sup>。

国内 MOOC 平台(13 个): ewant 育网开放教育平台(台湾)<sup>⑧</sup>、IBM 大数据大学<sup>⑨</sup>、爱课程网<sup>⑩</sup>、北大慕课<sup>⑪</sup>、好大学在线<sup>⑫</sup>、华文慕课<sup>⑬</sup>、清华大学磨课师(台湾)<sup>⑭</sup>、网易公开课<sup>⑮</sup>、网易云课堂<sup>⑯</sup>、新浪公开课<sup>⑰</sup>、学堂在线<sup>⑱</sup>、中国大学 MOOC<sup>⑲</sup>、中华开放教育平台(台湾)<sup>⑳</sup>。

截止至 2018 年 4 月,在上述 17 个平台上,笔者以“数据科学”或“data science”为检索词,获取原始的 MOOC 平台返回的课程列表,只选取课程标题或课程介绍中出现“数据科学”或“data science”字样的课程,最终获取数据科学课程 87 门。其中,除去平台间任

① <https://mooc.guokr.com/course/>

② <https://mooc.guokr.com/post/605193/>

③ 限于客观原因,部分国外 MOOC 平台在国内访问受限,因此不作为本文的研究对象,例如 Udemy(<https://www.udemy.com/>)响应时间过长。

④ <https://www.coursera.org/>

⑤ <https://www.edx.org/>

⑥ <https://ocw.mit.edu/index.htm>

⑦ <https://cn.udacity.com/>

⑧ <http://www.ewant.org/>

⑨ <http://bigdatauniversity.mooccollege.cn/>

⑩ <http://www.icourses.cn/>

⑪ <http://mooc.pku.edu.cn/>

⑫ <http://www.cnmooc.org/>

⑬ <http://www.chinesemooc.org/>

⑭ <http://mooc.nthu.edu.tw/>

⑮ <http://open.163.com/>

⑯ <http://study.163.com/>

⑰ <http://open.sina.com.cn/>

⑱ <http://www.xuetangx.com/>

⑲ <http://www.icourse163.org>

⑳ <https://www.openedu.tw/>

## 基于 MOOC 平台的国内外数据科学课程建设对比分析及经验启示

Lessons from the Comparison of Data Science in MOOC at Home and Abroad

桂思思 任珂 王玉琦 程齐凯

课教师及课程结构均相同的重复开设的课程 6 门<sup>①</sup>, 不重复开设的数据科学课程共计 81 门, 其中国外、国

内平台上开设的课程分别为 63 门和 18 门, 详细统计数据如表 1 所示。

表 1 MOOC 平台开设(全部)数据科学课程的数量统计表

| 国外 MOOC 平台         | 课程数量 | 国内 MOOC 平台 | 课程数量 |
|--------------------|------|------------|------|
| Coursera           | 27   | IBM 大数据大学  | 4    |
| edX                | 33   | 好大学在线      | 7    |
| MIT-OpenCourseWare | 2    | 网易公开课      | 1    |
| Udacity            | 1    | 网易云课堂      | 7    |
|                    |      | 学堂在线       | 5    |
| 总计                 | 63   | 总计         | 24   |

注:表中未列出的已调研的 MOOC 平台开设数据科学课程数量为 0。

#### 4 MOOC 平台数据科学课程开设情况分析

##### 4.1 平台维度

###### 4.1.1 平台建设

国内外的 MOOC 实践均积极构建多语种平台, 力图打破教学的语言隔阂。国外平台的相关解决方案已比较完善, 为课程相关视频配上英、中、日、西文字幕; 部分平台也开发中文界面, 如 Udacity, 用户可直接使用中文进行课程检索。国内正积极建立中、英双语平台, 例如, 学堂在线设立了单独的英文页面, 用以介绍平台相关情况, 好大学在线已开发了英文的平台界面, 但课程资源仍以中文为主。

另外, 国内外平台均针对不同用户(未注册、已登录未获取学习路径、已登录并已获得学习路径)分别设计不同界面, 因此不同平台之间界面结构相似, 用户在不同平台间切换时, 熟悉度高, 用户体验好。

###### 4.1.2 资源数量

开设了数据科学课程的国内外平台数目无明显差异, 而两类平台上的课程数量则差异明显。由表 1 可知, 本文调研的 17 个国内外平台中, 开设了数据科学课程的平台共计 9 个(占比超过 50%), 其中国外平台 4 个, 国内平台 5 个, 但是, 开设于 4 个国外平台上的课程数目(63 门)大约是开设在 5 个国内平台的课程数目的 2.5 倍(24 门)。剔除好大学在线上无法获

取上线日期的 2 门课程<sup>②</sup>, 5 个国内平台上开设的数据科学课程的数目随年份增长如图 2 所示。其中, 横轴表示年份, 纵轴为截止至该年年底平台上开设的数据科学课程总量。由图 2 可知, 在国内平台中, 自 2015 年首次开设数据科学课程以来, 除了网易公开课, 其他 4 个平台开设的数据科学课程数目均呈增长趋势, 且网易云课堂增长最快, 已开设的课程最多。

###### 4.1.3 资源质量

国外平台上无重复收录的数据科学课程, 国内平台却存在课程重复收录、原创资源过少的问题, 例如: ①平台内部存在重复课程资源: IBM 大数据大学分别开设了《数据科学方法论》的中文版和英文版课程, 除了授课语种, 其他课程设置完全一样; ②国内平台间存在重复课程资源: IBM 大数据大学开设的全部 4 门课程均重复收录于好大学在线; ③与国外平台存在重复课程资源: 学堂在线 2 门课程的链接自动重定位至 edX。该问题的产生一方面与平台定位相关, 例如网易公开课在鼓励用户发布原创课程的基础上, 网站运营方将收录其他平台的课程(国外知名大学公开课、可汗学院和 TED 等视频课程); 另一方面与课程资源配置相关, 我国 MOOC 平台应加强与国内各高校、各单位的合作, 吸引更多的优质资源。

①IBM 的 4 门课程均重复收录于好大学在线(数据科学方法论(中文版/英文版)、数据科学简介、数据科学入门、利用开源工具进行数据科学动手实践); edX 的 2 门课程重复收录于学堂在线(Data Science Essentials(数据科学与机器学习要领)、Data Science Ethics(数据科学道德规范))

②Scala for Data Science、[系列课]—数据科学家

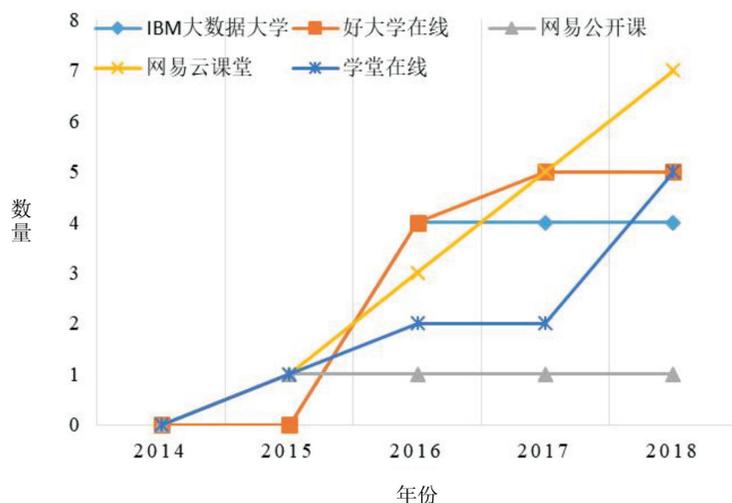


图2 国内MOOC平台开设课程数量变化趋势图

## 4.2 组织机制维度

### 4.2.1 开设单位

国内和国外平台上数据科学课程的开设单位分为几类:知名大学、其他教育机构/组织、行业组织或个人。知名大学或其他教育机构/组织开设的课程理论系统性强,能保证课程质量,但是相比较而言,知名

大学开设的课程等同于大学课程,部分可等效于知名大学的学位课程。行业组织开设的课程以职业为导向,提供前沿的课程,课程内容实用性强,可以应用于工作中。个人主要根据自身工作实践经验授课。国内外平台上的81门数据科学课程按照开设单位类别的分布如图3所示。

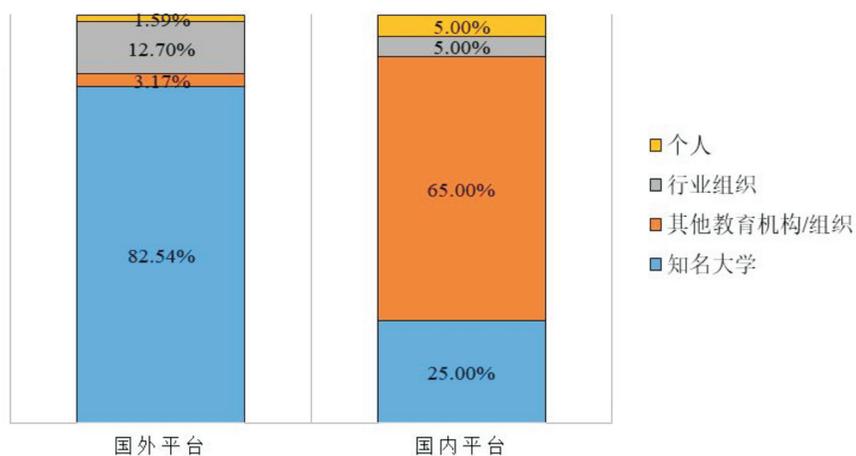


图3 国内外MOOC平台81门课程数量按开设单位类别的分布图

国外平台上数据科学课程的开设单位主要以知名大学与行业组织为主,分别占比82.54%和12.07%。开设数据科学课程的知名大学共计14个,平均每个大学开课3.7门,开课最多的学校分别为约翰

霍普金斯大学(15门)、哈佛大学(10门)、以及加利福尼亚大学伯克利分校(5门)。从国别来看,约70%的大学为美国的大学,其他大学为澳大利亚的阿德莱德大学、俄罗斯的国立高等经济大学、荷兰的埃因霍温

## 基于 MOOC 平台的国内外数据科学课程建设对比分析及经验启示

Lessons from the Comparison of Data Science in MOOC at Home and Abroad

桂思思 任珂 王玉琦 程齐凯

科技大学、墨西哥的墨西哥国立自治大学。上述 14 所大学中,3 所为 iSchools 联盟成员,表明数据科学与图书情报领域有一定联系,开展数据科学的研究不仅可刺激图书情报对自身理论和方法进行反思,还可激发图书情报的活力。

国内平台上数据科学课程的开设单位主要以其他教育机构/组织为主(IBM 大数据大学、WISE 学习天地、艾数教育、数盟大学堂、松鼠学堂、天善智能云课堂),知名大学开设的数据科学课程数量偏少。值得一提的是,部分国内知名大学启用联合授课模式,充分整合不同机构的资源,例如“计算思维和数据科学导论”由麻省理工学院、清华大学共同开设,数据科学引论(Python 之道)由上海交通大学与 IBM 大数据大学共同开设。

与国外相比,对于数据科学课程,国内知名大学虽尝试联合授课的模式,但是整体上开设课程数量的比例偏低,表明现今我国大学对数据科学教育的重视程度不足,一定程度上缺乏学科敏感度。

#### 4.2.2 课程教师

国内外课程的授课教师可分为实践教师和专业教师两类:①实践教师是个人组织者开设课程的主要形式,他们以个人的身份来组织课件,讲授课程,这些教师包括数据分析师、企业项目管理者 and 数据科学家等。他们将各自领域中的精华知识进行提取和整理,为学习者提供行之有效的实践经验;②专业教师是知名大学开设课程的主要形式,他们可以利用大学的资源对课程严格把关,提高课程质量,同时利用大学的知名度和影响力吸引学生。其中,超过 50% 的课程是由一名教师进行全部课程的讲授,以保证课程风格的统一性和连续性,例如华盛顿大学在 Coursera 开设的“Data Science at Scale”专项课程(4 门子课程);其他课程由多名教师共同授课,例如哥伦比亚大学开设的“Data Science for Executives”系列课程(3 门)由来自不同学科的 18 名教师参与授课。

### 4.3 课程结构维度

#### 4.3.1 课程类型

国内外平台上开设的课程类型有差异:国外平台上课程的主要类型主要为三类:常规性课程、专项课程和

系列课程(Xseries);国内平台上课程类型主要为五类:常规性课程、专项课程、系列课程(Xseries),国内自主模式及随堂模式。

常规性课程、随堂模式以及国内自主模式均为单一课程,而专项课程和系列课程是课程集合。具体如下:常规性课程与随堂模式类似,均专注于某个知识面所进行讲授的课程,模拟真实的线下课堂的教学类型,规定开课时间、课时,用户观看视频后需在规定的时间内完成指定的习题,并进行结课测试,还可以在平台论坛上和老师、同学进行交流互动。部分课程在学习者全程参与并达到测试及格线后,会提供结课证明或证书,十分接近真实的课堂体验。国内自主模式较随堂模式而言,没有时间的限制,不需要等待开课时间,用户可以随时加入课堂,并进行视频观看、内容讨论和课程测试等活动。专项课程将一系列常规性课程按照主题相似度结合起来,并根据行业变化市场需求,实时更新与知名企业合作的结业项目(Capstone Project)内容,以帮助学习者在实践中系统掌握知识,当学习者完成结业项目后,可以获得相应证书。系列课程提供了一种和学校课程设置相似的体系,但是不提供结业项目,只是对常规性课程进行有机整合。

国内外平台上的 81 门的数据科学课程按照课程类型的统计如表 2 所示。可以看出,国内、外课程均涉及了单一课程以及课程集合两种,且国内课程结合了我国实际情况,开设了 5 门国内自主模式课程、3 门随堂模式课程,然而国内课程在课程集合模式上数量稍显不足。

表 2 国内外平台上的 81 门数据科学课程按照课程类型的统计表

|        | 国内课程 | 国外课程 |
|--------|------|------|
| 系列课程   | 1    | 6    |
| 常规性课程  | 8    | 52   |
| 专项课程   | 1    | 5    |
| 国内自主模式 | 5    | —    |
| 随堂模式   | 3    | —    |
| 总计     | 18   | 63   |

#### 4.3.2 课程内容与主题

本文抽取 MOOC 课程的“课程简介”和“学习任务”板块的关键词,并根据数据生命周期理论(Digital Life

Cycle)<sup>[28]</sup>,将课程内容主题分为四个方面,具体如下:

(1)相关理论:包括数据科学相关概念、数据科学的法律和道德问题和数据科学家的职业前景。这方面课程着眼于数据科学开发和实际利用中可能会涉及到的学科理论、隐私和伦理问题,重点关注当今社会强调的团队合作意识,致力于培养学生良好的学术道德规范和卓越的团队管理才能。如“数据科学伦理(Data Science Ethics)”介绍了在数据科学研究中使用数据的知情同意权和数据的隐私保护问题。

(2)数据准备:包括数据处理和数据库利用。良好的数据处理会大大减小数据分析等后续工作的难度,并在一定程度上提高结果分析的准确度。如约翰霍普金斯大学“数据科学(Data Science)”专项课程中的子课程“数据的获取与处理(Getting and Cleaning Data)”详细阐述了数据的获取以及数据清洗等一系列数据前期准备工作。

(3)数据统计分析:包括机器学习中对模型的搭建、运用编程语言或工具对数据进行统计分析。如数据科学与分析系列课程(Data Science and Analytics Xseries)详细介绍了运用统计思想和条件概率,以及机器学习和算法,对数据进行统计分析的基本技能。

(4)数据可视化:研究数据视觉表现形式,是现在非常流行和有效的展示自我成果的方式。华盛顿大学“大规模数据科学(Data Science at Scale)”专项课程中的子课程“发布研究成果:形象化、伦理和再现性”,详细阐述了发布研究成果时结果形象化的重要意义,强调数据可视化可以更加清晰地展示数据所表示的含义。

总体来说,数据科学课程主要可分为理论性和实践性两个方面,除第一部分为理论性课程外,其余三部分均为实践性课程。讲授理论内容的课程数量较少,但其作为数据科学的基础性内容,对于了解和掌握数据科学的基本知识起着奠基性作用,实践课程作为数据科学的核心,在数据科学的利用中起着决定性作用,两者相辅相成,共同构筑数据科学的系统性知识框架。

#### 4.3.3 课程级别定位

国内外平台上的81门数据课程的级别均分为3类(初级、中级和通识)。通识课是指学习者应掌握的

基础教育课程;初级课程主要针对计算机科学、生物或其他专业的在校生,或对数据科学的前沿领域感兴趣的学习者,对背景知识和相关实践经验的要求很少或基本不要求,比如约翰霍普金斯大学开设的“数据科学(Data Science)”课程,主要对数据科学的理论和应用进行基础性介绍;中级课程相较于初级课程而言难度加大,需要学习者具有特定的知识背景和实践经验,如约翰霍普金斯大学开设的“基因组数据科学(Genomic Data Science)”课程要求学习者具有医学或生物学领域的知识,并要求掌握一定的医学统计学技能。

然而国内外平台上不同级别课程开设比例不同,具体见图4。国内和国外平台上通识课程开设初级课程的开设比例相差不大(<10%),然而在中级课程的开设比例相差较大(>15%):国外平台上有15门中级课程,而国内平台上仅1门中级课程(“利用开源工具进行数据科学动手实践”),国内平台需考虑设置一定数量的中级课程,加大内容难度、加深知识深度、提升课程质量。

#### 4.3.4 课程资源展示

课程资源属于课程丰富性和课程多样性的直接展示方式。本文根据“课程简介”、“讨论”和“反馈”等版块归纳出3种常见课程资源呈现方式:授课视频、案例讨论和学习反馈,其中学习反馈包括随堂测验、综合测验、课后作业和作业互评。

对于课程资源呈现方式,国外的课程对以上3种方式均有涉及且实现效果良好,而国内MOOC课程主要以视频形式为主,在案例讨论与学习反馈设置上稍显不足,28%课程缺乏案例讨论环节、33%课程缺乏实验反馈环节。因此我国应加强课程资源呈现方式的建设,为学习者提供更加多样的学习途径,使他们可以更加有针对性地填补自己知识的空缺。

#### 4.3.5 课程交互系统

良好的交互是现代MOOC平台区别于以往单纯课程视频集合的最主要特征之一。通过交互系统,学习者可以和同伴或老师进行互动交流,发现和解决学习上遇到的问题,更好地进行课程学习,交互方式主要有以下几类:

## 基于 MOOC 平台的国内外数据科学课程建设对比分析及经验启示

Lessons from the Comparison of Data Science in MOOC at Home and Abroad

桂思思 任珂 王玉琦 程齐凯

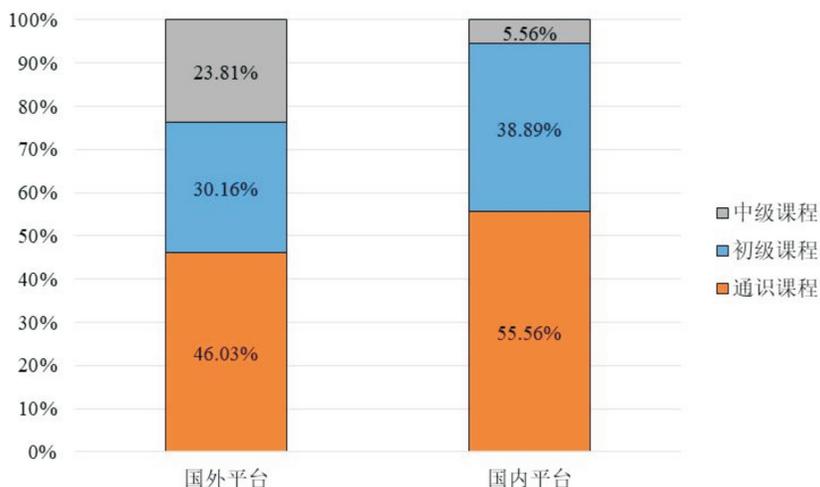


图 4 国内外 MOOC 平台 81 门课程数量按课程级别定位的分布图

(1) 电子邮件。已调研的课程均使用了电子邮件,当学习者注册成功一门数据科学课程后,系统自动发送课程相关邮件,并及时督促学习者按时上课。

(2) 社交网络平台。学习者通过社交网络了解最新课程动态,并提出自己的问题与老师探讨。国外 MOOC 平台支持 LinkedIn、Twitter、Google +、YouTube、Facebook 等,同时为了方便我国用户的使用,也增加了微信平台的社交互动。国内 MOOC 平台主要包括新浪微博、豆瓣、QQ 空间、人人网等。

(3) 课程论坛。学习者在论坛中通过发帖提出问题,通过回帖与同学交流。课程论坛可被细分,比如网易云课堂开设的“解密数据科学”课程将论坛分为

周论坛和总论坛,为不同时间段学习的用户提供更加有针对性的交流平台。

通过交互系统的作用,可以使学习者的问题得到及时解答,查漏补缺,同时可以看到,交互系统也针对不同地区学习者的使用习惯推出不同的交互方式,以方便用户使用,但总体差别不大。

## 5 对我国 MOOC 数据科学教育的启示

根据平台—组织机制—课程结构三维 MOOC 课程分析框架,表 3 对 MOOC 数据科学课程开设特点进行汇总,并在此基础上对我国 MOOC 数据科学教育提出启示。

表 3 MOOC 数据科学课程开设特点

| 维度   | 特点   |
|------|--|
| 平台   | 平台建设:国外平台体系成熟;国内正处于发展初期<br>资源数量:开设课程数量上具有一定差异,但是国内数据科学 MOOC 课程具有良好的发展势头<br>资源质量:国外平台为原创课程,国内平台上存在少数重复收录的课程   |
| 组织机制 | 开设单位:类别相同,但比例差异大,国外为知名大学及行业机构;国内授课的主体主要为其他教育机构/组织,其次为知名大学<br>课程教师:均包含实践教师和专业教师两类   |
| 课程结构 | 课程类型:国外以常规性课程、专项课程和 Xseries 三类为主;国内课程在上述类型基础上,还包含了随堂模式以及国内自主模式<br>课程内容与主题:相似,均包括相关理论、数据准备、数据统计分析、数据可视化<br>课程级别:均设有初级、中级和通识,然而国内平台的中级课程开设比例较少<br>课程资源展示方式:相似,均包括授课视频、案例讨论和学习反馈三类<br>课程交互系统:相似,均包括电子邮件、社交网络论坛和课程论坛三类 |

### 5.1 平台维度——加强平台与各教育机构的合作,增加原创课程数量

国外 MOOC 平台的建设,从 2011 年发展至今已日趋成熟,已形成自身的学习和管理模式。虽然我国 MOOC 平台起步晚于国外,但是具有良好的发展趋势,例如开设了数据课程的国内平台数目与国外平台数据基本相同,开设的科学数据的课程数量呈逐年上升趋势,平台之间建立合作关系等。然而国内平台上的原创课程资源相对匮乏,例如平台间存在重复开设的课程,不利于我国数据科学的发展和建设。因此我国应该大力支持优质的 MOOC 平台,鼓励各教育机构积极进行网络课程的制作,使两者形成良好且长期的合作,产出更多的优质资源,丰富 MOOC 学习平台,推动数据科学等学科的发展。

### 5.2 组织机制维度——丰富数据科学课程来源,保证课程质量

目前国外 MOOC 平台合作最多的是国外知名大学,其次为行业机构,在确保课程质量和理论体系的完整性的同时,也联系实际以及市场需求,结合实际工作中的经验更具有针对性地进行授课。需要说明的是,由知名大学开设的部分课程也联合企业共同授课,依托企业中实际的项目需求,让学习者在实践中灵活运用知识。而国内 MOOC 平台课程则主要由其他教育机构/组织开设,其次为知名大学开设,再次为行业机构。虽然部分大学已尝试联合国外大学或者国内行业机构共同授课,但是由大学开设的课程数量偏低,难以保证课程质量及体系的完整性;同时行业机构开设课程数量偏低,易导致课程实用性低。因此,国内 MOOC 平台一方面需要加强与知名大学之间的合作,例如 iSchool 联盟成员等,丰富以大学为课程提供者的数据科学课程,使更多的学习者能够接触到优质的教育资源,另一方面,需要加强与行业机构间的合作,丰富具有职业导向的课程,使课程开设目标更贴合市场需求,提升课程的实用性。

### 5.3 课程结构维度——建立完善的 MOOC 数据科学课程体系

虽然我国数据科学教育处于发展初期,但是已经逐步走上了正轨,例如国内平台上课程在课程主题、

课程级别、资源展示形式、课程交互形式、课程级别的设置上基本与国外平台上课程相同。除此之外,课程类型设置更结合实际情况,在国外基础上(常规性课程、专项课程、Xseries 系列课程)加入随堂模式以及国内自主模式。即便如此,国内平台上数据科学内容建设上还有一定的改进空间,例如:①课程资源展示形式主要以视频形式为主,需增加案例讨论和学习反馈的数量;②中级数据科学课程数量偏少,需增加课程的广度与深度,从而提高中级课程的比例;③课程内容的设置可依据我国实际情况,根据我国学习者的学习习惯与需求,建立有我国特色的、完善的 MOOC 数据科学课程体系。

### 5.4 其他

数据科学是一门研究从数据中提取知识的学科,因此应加强其与各专业知识的结合。在今天,对于这些庞大且不断增长的数据,专业知识将是必不可少的。在 81 门课程中,只有少数有关数据科学运用到专业领域中的课程,例如“基因组数据科学(Genomic Data Science)”,其他课程均没有涉及具体的领域。情报学是一门具有极强的数据敏感性的学科,致力于从数据中挖掘领域的研究热点和前沿动态,因此将数据科学和情报学结合可以推动两个学科共同发展。

## 6 结语

数据科学是当今信息社会对数据利用的必然需求。本文通过国内外对比分析,以平台、组织机制和课程结构三个维度为切入点,围绕不同 10 个要素对 17 个平台上的 81 门数据科学课程进行全面分析,深入了解其课程模式和办学模式。研究表明,虽然我国数据科学教育处于起步阶段,但是发展趋势较好,在平台、组织机构以及课程结构等方面仍具有一定的改善空间,例如加强平台与各教育机构的合作以增加原创课程的数量,丰富课程来源以保证课程质量等。本研究希望通过建立完善的 MOOC 数据科学课程体系,最终丰富我国 MOOC 数据科学教育,培养社会需要的数据科学家,辅助企业和政府进行决策支持。未来我们将从面授课程和 MOOC 课程的对比分析出发,进一步完善数据科学教育对比分析。

## 基于 MOOC 平台的国内外数据科学课程建设对比分析及经验启示

Lessons from the Comparison of Data Science in MOOC at Home and Abroad

桂思思 任珂 王玉琦 程齐凯

## 参考文献

- Hollands F M, Tirthali D. MOOCs: Expectations and Reality (Full report) [EB/OL]. [2018-04-18]. <https://oerknowledgecloud.org/content/moocs-expectations-and-reality-full-report>.
- 26 王颖, 张金磊, 张宝辉. 大规模网络开放课程 (MOOC) 典型项目特征分析及启示[J]. 远程教育杂志, 2013(4): 67-75.
- 3 朱玉奴. 国外数据素质教育及图书馆服务模式研究 [J]. 图书馆情报工作, 2014, 58(5): 29-34.
- 4,8 叶鹰, 马费成. 数据科学兴起及其与信息科学的关联 [J]. 情报学报, 2015, 34(6): 575-580.
- 5 Patil T, Davenport T. Data Scientist: the Sexiest Job of the 21st Century[J]. Harvard Business Review, 2012.
- 6 Manyika J, Chui M, Brown B, et al. Big Data: The Next Frontier for Innovation, Competition, and Productivity[EB/OL]. [2018-03-23]. <http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation>.
- 7 Dhar V. Data Science and Prediction[J]. Communications of the ACM, 2013, 56(12): 64-73.
- 9 Hazen B T, Boone C A, Ezell J D, et al. Data Quality for Data Science, Predictive Analytics, and Big Data in Supply Chain Management: An Introduction to the Problem and Suggestions for Research and Applications[J]. International Journal of Production Economics, 2014, 154: 72-80.
- 10 许荻迪, 江小蕾. 数据科学的发展及其在客户细分中的应用 [J]. 商业经济, 2015(7): 84-88.
- 11 王迪, 王东雨. 美国数据科学课程设置对信息素养的影响研究[J]. 商, 2016(14): 200-200.
- 12 苏日娜, 肖鹏, 林毅鸿. 美国数据科学课程设置对信息素养的影响研究[J/OL]. 图书馆论坛, 2018(10). [2018-04-17]. <http://kns.cnki.net/kcms/detail/44.1306.G2.20170329.0121.002.html>.
- 13 徐昊, 秦玥, 黄岚. 面向通识教育的数据科学课程建设[J]. 计算机教育, 2016(8): 158-162.
- 14 Burd E L, Smith S P, Reisman S. Exploring Business Models for MOOCs in Higher Education[J]. Innovative Higher Education, 2015, 40(1): 37-49.
- 15 Li Y, Powell S. MOOCs and Open Education: Implications for Higher Education [EB/OL]. 2013. [2018-03-26]. [https://s3.amazonaws.com/academia.edu/documents/33781834/MOOCs.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1523958395&Signature=PjFXjaQOpdlhE0pV%2Bt0pe2cfFc%3D&response-content-disposition=inline%3B%20filename%3DThis\\_file\\_I\\_got\\_from\\_the\\_net\\_its\\_very\\_im.pdf](https://s3.amazonaws.com/academia.edu/documents/33781834/MOOCs.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1523958395&Signature=PjFXjaQOpdlhE0pV%2Bt0pe2cfFc%3D&response-content-disposition=inline%3B%20filename%3DThis_file_I_got_from_the_net_its_very_im.pdf).
- 16 Wolff A, Hudson L, Kortuem G. The Role of Data Literacy within a MOOC Analysis[C] //LAK '16: The 6th International Learning Analytics and Knowledge Conference. 2016.
- 17 Bond P. Information Literacy in MOOCs[J]. Current Issues in Emerging eLearning, 2015, 2(1): 6.
- 18 Hardin J, Hoerl R, Horton N J, et al. Data Science in Statistics Curricula: Preparing Students to "Think with Data"[J]. The American Statistician, 2015, 69(4): 343-353.
- 19 Liyanagunawardena T R. Massive Open Online Courses[J/OL]. Humanities, 2015, 4(1). [2018-04-18]. <http://www.mdpi.com/2076-0787/4/1/35>.
- 20 Christensen G, Steinmetz A, Alcorn B, et al. The MOOC Phenomenon: Who Takes Massive Open Online Courses and Why? [J/OL]. Social Science Electronic Publishing, 2014. [2018-04-18]. [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2350964](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2350964).
- 21 Andersen R K, Garp K, Nellemann K, et al. A Framework for the Analysis of Collaborative and Interactive Elements in MOOCs [C] //European Conference on e-Learning. 2014: 19.
- 22 Chauhan J, Goel A. An Analysis of Video Lecture in MOOC[C] //11th International Conference on Information and Communication Technologies in Education, Research, and Industrial Applications: Integration, Harmonization, and Knowledge Transfer, Lviv, Ukraine, 2015: 35-50.
- 23 Mukala P, Buijs J C, Leemans M, et al. Learning Analytics on Coursera Event Data: A Process Mining Approach[C] //SIMPDA: 5th International Symposium on Data-Driven Process Discovery and Analysis. 2015: 18-32.
- 24 Veeramachaneni K, O'Reilly U-M, Taylor C. Towards Feature Engineering at Scale for Data from Massive Open Online Courses [EB/OL]. [2018-04-18]. <https://arxiv.org/abs/1407.5238>.
- 25 Williams K. Content Analysis of Coursera, edX, and Udacity Course Platforms [D]. Charlottetown, Canada: University of Prince Edward Island, 2014.
- 27 Meyer R, Gaskill M, Vu P. Rating User Interface and Universal Instructional Design in MOOC Course Design[J]. Revue internationale des technologies en pédagogie universitaire/International Journal of Technologies in Higher Education, 2015, 12(1): 62-74.
- 28 Yu X, Wen Q. A View about Cloud Data Security from Data Life Cycle[C] //CiSE 2010: International Conference on Computational Intelligence and Software Engineering, IEEE, 2010: 1-4.

(收稿日期: 2017-03-02)